TRANSITION MODEL FOR CORONA VIRUS MANAGEMENT

A. Djogbenou [†], C. Gourieroux [*], J. Jasiak [‡], P. Rilstone [§] and M. Bandehali [¶]

first version: May 14, 2020
revised: June 12, 2020

Confidential and Preliminary Version

This paper summarizes the outcomes of research on the IPHIS-REPORT.csv dataset provided by the Ontario Ministry of Health on May 07, 2020.
The analysis is focused on the management of various stages of medical treatment of patients infected by COVID-19 in Ontario. We trace out daily transitions of infected individuals between different types of medical care and account for their sojourn times to reveal duration dependence. The transition model is used to estimate and predict the counts of patients treated for COVID-19, while accommodating the truncation and right censoring in the sample.

Keywords: Covid-19, Transition Model, Medical Care, Multinomial Logit, Duration Dependence, Right Truncation.

[†]York University, Canada, *e-mail*: daa@yorku.ca
[*]University of Toronto, Toulouse School of Economics and CREST,*e-mail*: gouriero@ensae.fr
[‡]York University, Canada, *e-mail*: jasiakj@yorku.ca
[§]York University, Canada, *e-mail*: pril@yorku.ca
[¶]York University, Canada, *e-mail*: mbandehali@yorku.ca
The authors thank the MOH for providing the data.

1 Introduction

A basic epidemiological model SIRD distinguishes 4 individual states [see e.g. Vinnicky, White (2010), Yan, Chowel (2019), Toda (2020)]. State S of susceptible comprises individuals who are healthy and not yet immunized. In state I of infected, the individuals are infected and not yet recovered. An infected individual can either recover, i.e. move to state R, or decease, i.e. move to state D. This causal scheme, which can be summarized as follows:

$$S \to I \begin{cases} \nearrow R \\ \searrow D \end{cases}$$

is a chain that involves two episodes. The first episode between states S and I concerns the propagation phase of the disease and the process of detection of infected individuals. The second episode concerns the disease monitoring that follows a diagnosed infection. In this paper, we are interested in the analysis of individual medical care histories. Therefore, we divide the state of Infected into states of medical care of increasing intensity, which depend on the severity of symptoms. These additional states include Hospitalization, Intensive Care Unit (ICU), Ventilation and Intubation.

Our analysis of COVID-19 infections in Ontario is based on daily records of 18722 individuals who were diagnosed with COVID-19 over the period of 104 days between January 23 and May 05, 2020 and reported in the database IPHIS_REPORT.csv by the iPHIS (integrated Public Health Information System) and CORES (Toronto Public Health Coronavirus Rapid Entry System).

The objective of our research is to introduce a modelling approach for the analysis of counts of patients under medical care that produces reasonably accurate results, given the complexity of problems encountered in the IPHIS_REPORT.csv dataset. Some of those problems are related to the data collection method. These are, for example, the missing and misreported dates of medical treatments, the recovery dates being unavailable, or unknown outcomes of medical care. Other problems are related to the statistical analysis of data with right censoring and truncation, which is amplified by the delay in data reporting of up to 10 last days of the sampling period.

The publicly available data on COVID-19 infections in Ontario are aggregated at various levels. For example, let us consider the data provided by the Public Health Ontario

(PHO) [see, PHO (2020 a), (2020 b)]. These sources provide accounts of confirmed, and deceased individuals by age or region. The Enhanced Epidemiological Summary [PHO 2020 a)] shows additional counts of patients who are hospitalized and in ICU. Below, we display similar series of counts computed from our sample until May 04, after a preliminary individual data adjustment for misreported data and outliers [see Appendix A.2]. Figures 1 and 2 show the cumulated and daily counts of Diagnosed and Deceased individuals, respectively.

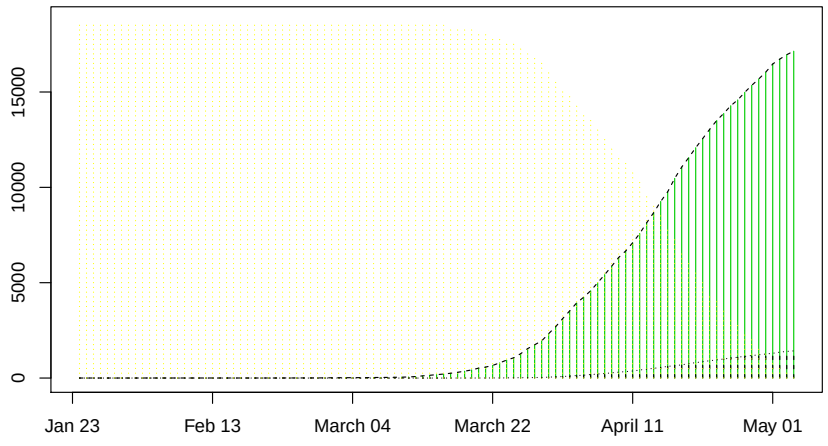


Figure 1: Counts of Undiagnosed, Diagnosed and Deceased

These figures can be compared with Figures 1 and 4 in PHO (2020 b, pages 3 and 6) that report the confirmed cases and deaths over a longer sampling period ending on May 27. The patterns prior to May 05 revealed in Figures 1 and 2 above are close to those displayed in Figures 1 and 4 of PHO (2020 b). Note that due to a reporting lag, the last few daily counts in Figures 1 and 2 need to be considered with caution.

In Figure 2, the daily increments reveal hump-shaped patterns, which resemble the curves of standard epidemiological SIRD models. The dates of peaks are about April

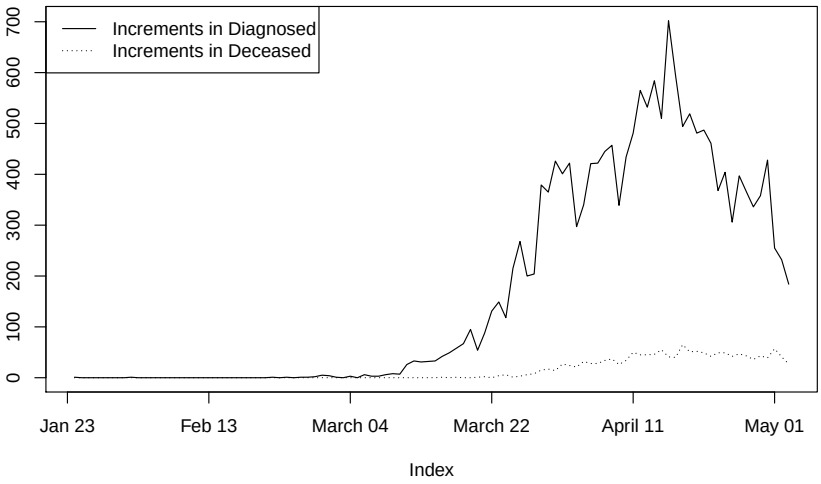


Figure 2: Daily Counts of Diagnosed and Deceased

15 for the Diagnosed and May 01 for the Deceased. There is a delay of about 2 weeks between the two peaks due to the length of medical treatment of COVID-19. There also is a seasonal effect in reporting with periodicity of 7 days that creates a jagged pattern in Figure 2.

The curves of daily counts display a non-stationary behaviour with a phase of growth followed by a decrease. The aggregated data do not allow us to disentangle the possibly multiple sources of this non-stationarity, such as non-stationary propagation of the epidemic with an exponential increase in the early phase, the effect of disease detection that depends on the reliability and availability of tests, and the treatment for COVID-19. The latter one can improve over time due to better knowledge of the disease or worsen due to shortages of medical staff and equipment.

Figure 3 below shows the daily counts of patients under medical care and can be compared with Figure 1 (ICU, Hospitalization) in PHO (2020 a, page 3) where the sampling period is shorter and ends in April 22 [or with Figure 2 in PHO (2020 c)]. We observe that the peaks depend on the sampling period. They seem to be on about April 1 in the PHO (2020 a) and about April 11 in Figure 3. This is likely due to a kind of truncation bias. This bias is discussed later in Section 2.3 and adjusted for by considering the state

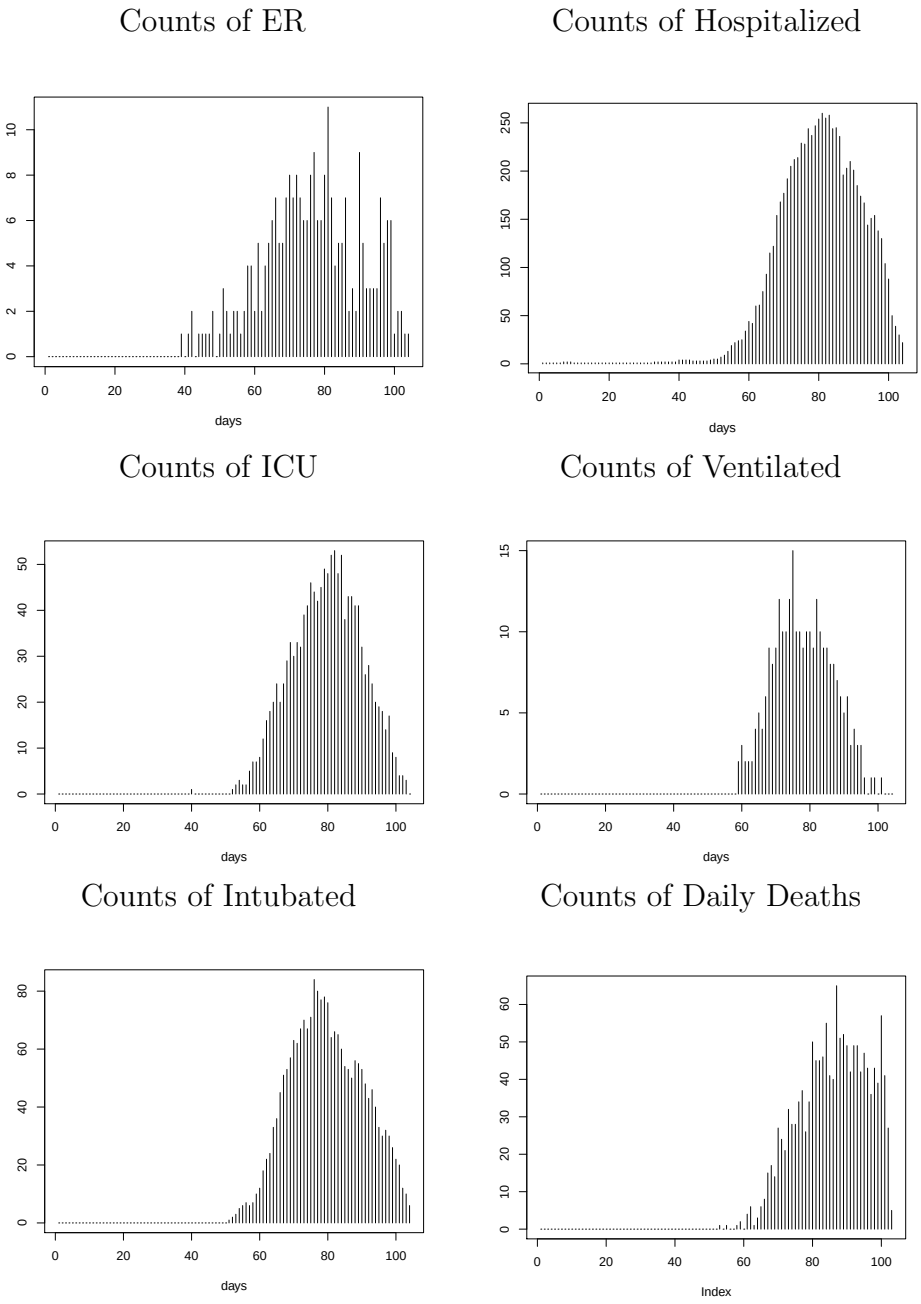


Fig 3. Counts of Medical Care

transitions instead of "crude" counts.

Figure 4 presents changes in daily counts displayed in Figure 3. It shows patterns that are different from those in Figure 3 in Aguerragibiria et.al. (2020), which illustrates observations over a period ending on April 22. That difference seemingly confirms a

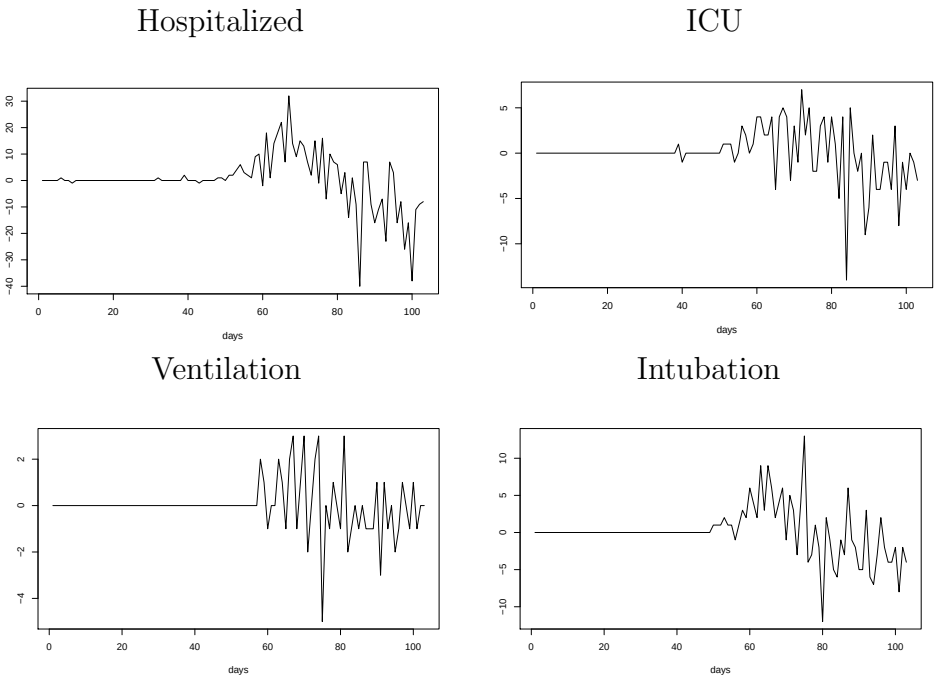


Fig 4. Changes in Daily Counts of Medical Care

misleading effect of truncation. Nevertheless, both Figures reveal daily fluctuations due to the fact that the counts of individuals under medical care are determined by both the entry and exit effects.

Our approach to examining the individual records relies on a transition model, where each individual is represented by a "history" variable, i.e. a sequence of states [see, e.g. Gourieroux, Jasiak (2007), Chapter 8]. It allows us to adjust for right censoring and attenuate the truncation bias. The state dynamics are defined through transition matrices, which are functions of the current and past environment.

An advantage of the transition model is that it allows us for including time and individual dependent explanatory variables, such as the time already spent in the state, i.e. the duration dependence.

The analysis of the transition model with duration dependence leads to the following observations:

The longer an individual stays hospitalized, the lower the probability of moving to ICU, or of being intubated. The probabilities of recovering and of dying of COVID-19 increase with the length of medical treatment. The probability of death seems to increase

with time spent in all types of medical care, at rates depending on the intensity of care.

The results concerning the probability to recover are affected by the lack of the date of recovery and need to be interpreted with caution.

The paper is organized as follows. Section 2 describes the states of medical care and illustrates the individual histories of transitions between states. The truncation problem is discussed too. Section 3 presents the empirical data analysis and summary statistics computed under a simplifying assumption of a homogeneous Markov chain. Then, we discuss the stationarity assumption of the medical care process. Section 4 presents the transition model with duration dependence, where the probabilities of transition from each state depend on the time already spent in that state. Section 5 concludes. The IPHIS data set is described in Appendix A.1. The data adjustments and data reporting problems encountered in the dataset are outlined in Appendix 2. The summary statistics of sojourn times in each state are given in Appendix A.3. The supplementary figures are given in Appendix A.4.

2 The Discrete Time Transition Model

This Section introduces the states, the transitions and the discrete time dynamic transition model based on the individual

2.1 Individual Histories

Below, we provide a few examples of individual histories found in the data set and explain how they were transformed for our analysis.

1. Case Reported April 17, Recovered by May 05.
2. Case Reported March 29, in ER on March 26, Hospitalized April 03, ICU and Ventilation from April 6 until April 23, not Recovered.
3. Case reported April 16, Hospitalized on April 24, not Recovered until May 05.
4. Case reported April 02, Hospitalized March 31, in ICU April 02, Died April 03.
5. Case reported April 24, not Recovered.

These examples of individual histories of patients underlie our approach and suggest the selection of states given below.

2.2 States

The columns of the dataset given in Appendix A.1 allow us to distinguish 9 latent states. The states are redefined for compatibility with the transition model and the data availability.

State 0. Undiagnosed: from the beginning of the sampling period until the date in column M (case reported date), or a transition date to the next state. On "the case reported date" the individual is diagnosed COVID-19 infected and enters a follow-up that includes self-isolation and/or states of medical treatment.

State 1. D: Domiciled: from the case reported date until the medical care, death or recovery. As the population of Ontario was supposed to self-isolate during the sampling period, we include in D individuals who are not currently hospitalized and not yet recovered: see below the definition of recovered. Therefore we have D = ISO + Self-ISO. The isolation can be prior to, or after, a medical treatment. These will be denoted by D^1 and D^2, respectively. Columns AK and AL are supposed to contain the dates of isolation. However, there are very few cases of isolation reported and their locations are unknown.

State 2. ER: Emergency Room between dates in columns AE and AF.

State 3. H: Hospitalized due to Covid (hospitalization for other reasons is disregarded) between dates Y and Z.

State 4. ICU: Intensive Care Unit between dates AB and AC.

State 5. V: Ventilation between dates AN and AO.

State 6. T: Intubation between dates AH and AI.

To help adjust breathing, a machine is used to move air in and out of the patient lungs. Ventilators (also called respirators) are machines of different types, including computerized microprocessor controlled machines, as well as CPAC (Continuous Positive Airway Pressure) and non-invasive ventilators. This state of medical care is called ventilation. The state intubation refers to placing a tube in the patient's throat to help move air in and out of the lungs while protecting the airway, which is a long term ventilator dependence with a tracheotomy cannula.

In many cases, the time spent in hospitalization, intensive care, intubation, or under a ventilator overlap. Given that the states should be disjointed events, we proceed as follows: We treat an individual as being intubated if this stage overlaps with others

(because being intubated means that the individual is already in intensive care, under a ventilator and breathing through a tube). Similarly, if a patient is in intensive care and on a ventilator, we consider him/her to be using a ventilator. When a patient is in intensive care and under hospitalization, we consider him/her as in intensive care.

This distinct states do not obey the relationship $H \supset ICU \supset V \supset T$, and the differences between the states considered, marked with a star and the traditional counts given in PHO reports are as follows:

$$\begin{cases} T^* = T \\ V^* = V - T \\ ICU^* = ICU - V \\ H^* = H - ICU \end{cases} = \begin{cases} T = T^* \\ V = V^* + T^* \\ ICU = ICU^* + V^* + T^* \\ H = H^* + ICU^* + V^* + T^* \end{cases}$$

For ease of exposition, the non-starred symbols are used, assuming the reader is aware of the above distinction.

Accordingly, among the 18722 individuals diagnosed in our data set, 2047 individuals went through medical care for COVID-19 during the sampling period of 104 days. The total counts of individuals who at some point over the sampling period were in the distinct (starred) states of medical care given above, are: 138 in ER, 1376 in H (Hospitalization), 243 in ICU, 46 in V (Ventilation) and 244 in T (Intubation).

State 7. R: Recovered.

This state is an important component of a SIRD model. However, for COVID-19, the notion of recovery is not clearly defined and the date of recovery is unknown. The database has two sources, which are the iPHIS and CORES that are seemingly not reporting the recovery in the same way. Therefore, we have the "outcomes" in column AQ, and "resolved" in column AR, which are reported without the dates. In the current PHO practice, these have been combined with a date chosen in a conventional way. More precisely, in the PHO documents "the following cases are considered resolved" (but referred to as "recovered" in the daily Ontario provincial report for the media):

-cases that are reported as "recovered" in iPHIS based on the local public health unit assessment;

-cases that are not hospitalized and are 14 days past their symptoms onset date, or specimen collection date;

-cases that are currently hospitalized, have a case status of "closed" indicating that public health follow up is complete and are 14 days past their symptoms;

There is a lack of coherency in the definition, as several individuals reported "recovered" by the iPHIS may still have COVID-19 symptoms, and the definition depends on the local health unit.

In our analysis, we have adopted a similar approach to defining the state R at a conventional time. For that reason, some results concerning the time to recover have to be considered with caution.

State 8. DE: Deceased on date AS.

Individuals who are not recovered or deceased can remain in their last state on May 05, causing a right censoring problem.

2.3 Truncation

Let us now consider the state of all individuals on May 05. On that day, there are 13218 recovered individuals and 1429 deceased. Therefore there are 4075 right censored histories. on May 05. Among those 4046 are in D, 1 in ER, 22 Hospitalized and 6 Intubated. Due to the lack of a clear definition of recovery, there are 4046 individuals in state D who have been diagnosed since less than 14 days.

Let us now discuss the right truncation of individual histories for individuals who are under medical care on May 05. On May 05, in Ontario [see, Global News May 05, 2020] 1043 individuals are hospitalized. Among them 223 are in ICU and 166 are under ventilation (PHO definition). At the same time, 3504 critical care beds are available, including 2811 beds equipped with ventilators [see News, Office of the Premier, April 16, 2020, 1:00 pm]. Therefore, the IPHIS_REPORT.csv database suffers from a truncation problem as very few individuals under medical care on May 05 are reported.

As individuals in medical treatment may stay even for 3 weeks in medical care the counts given in Figure 3 are reliable up to day 80 - 85. In particular, the decreasing counts observed in Figure 3 are misleading and essentially due to this truncation in the data base.

In fact, the data given on May 05 show that the curves need to be projected upward to obtain correct predictions for May 05 and the following days [see, Section 4.3]. Similar

biases may appear also in an analysis of crude sojourn times in each state [see, e.g. Lapidus et al. (2020) for application to COVID-19 and Blanhaps et al. (2013), Schweer, Wickelhaus (2015) for theoretical results].

2.4 Specification of transition probabilities

The transitions of interest are conditional on being diagnosed. The following conditional state transitions are assumed to occur :
From D $\to$ D, or D $\to$ ER, or D $\to$ H or D$\to$ ICU, or D $\to$ V, or D $\to$ T, or D $\to$ R, or D $\to$ DE .
From ER $\to$ D, or ER $\to$ ER, or ER $\to$ H, or ER $\to$ R, or ER $\to$ DE.
From H $\to$ D, or H $\to$ ER, or H $\to$ H, or H $\to$ ICU, or H $\to$ V, or H $\to$ T, or H $\to$ R or H $\to$ DE.
From ICU $\to$ D, or ICU $\to$ H, or ICU $\to$ ICU, or ICU $\to$ V or ICU $\to$ T, or ICU $\to$ R, or ICU $\to$ DE .
From V $\to$ D, or V $\to$ H, or V $\to$ ICU or or V $\to$ V, or V $\to$ T, or V $\to$ DE.
From T $\to$ D, or T $\to$ H, or T $\to$ ICU, or T $\to$ V, or T $\to$ T, or T $\to$ R, or T $\to$ DE.
From R $\to$ R :absorbing state.
From DE $\to$ DE :absorbing state.

During a medical care episode some transitions are known to have probability 0. For example, an individual cannot go back to ICU after recovery, etc.

2.5 Joint Distribution of Individual Histories

The probabilistic model concerns the histories of diagnosed individuals over the set of states given above. The individuals are indexed by i, i=1,..,n, where n is the total size of the sample. The model is applied to the individual histories starting from a common date t_0 equal to the first detection date of January 23.

As the individuals are diagnosed at different dates, we introduce the state 0 of Undiagnosed in order to align the individual histories on a common time scale. The histories can be characterized either:

i) by the dates of jumps from one state to another, or equivalently by the ordered sequence of occupied states with given sojourn times in each state (this is how the examples

of individual histories in Section 2.1 are reported),

ii) or, by a qualitative time series with given occupied state indicators on each day. For example, the individual history:
00011111111111111111112333333444441111111111111111111111117
represents an individual, undiagnosed for 3 days of the sampling period, isolated (state D) for 18 days, who enters the ER and then remains in hospital for 6 days. Next, the patient is moved to ICU for 5 days, returns to state D and recovers.

These two representations are equivalent and their use depends on the purpose of a study. The IPHIS_REPORT.csv relies on representation i) which takes less space, while a dynamic model can be applied to either one representation. Our estimation method is based on the qualitative time series representation b) in order to accommodate easily the right censoring.

Let us denote by $\{Y_{i,t}, t = t_0, t_0 + 1, ..\}, i = 1, ..., n$ the individual histories since the common date t_0. At the beginning of the history, the individual is still undetected and undiagnosed with $Y_{i,t} = 0$ by convention. After diagnosis date $S_{0,i}$ that depends on each individual, $Y_{i,t}$ is a qualitative variable with 8 possible states described above. The main probabilistic assumption is the following:

Assumption A.1

The individual histories $\{S_{0,i}, Y_{i,t},\ t \geq S_{0,i}\}, i = 1, ..., n$ are independent.

As we focus our analysis on the medical care history, conditional on the diagnosis, only the conditional distribution of $\{Y_{i,t}, t \geq S_{0,i}\}$ given $S_{0,i}, Y_{i,S_{0,i}}$ is relevant. This conditional distribution is characterized by a sequence of transition matrices $P_{i,t}$ that provides the conditional distributions of $Y_{i,t}$ given $S_{0,i}, \underline{Y_{i,t-1}}$, where $\underline{Y_{i,t-1}}$ contains all past observations up to and including time $t - 1$.

Assumption A.1 allows for a variety of specifications for the sequence of transition matrices, some of which are examined in the empirical analysis in Sections 3 and 4.

i) **Homogeneous Markov Chain**

Assumption A.2 a)

$P_{i,t} = P$, independent of i and t, $i = 1, ..., n$, $t > S_{0,i}$.

This assumption can be equivalently written under representation i) in terms of dates of jumps and states following those jumps. This representation is denoted by:

$\{S_{l,i}, \tilde{Y}_{l,i} = Y_{S_{l,i},i},\ l = 0, 1, ...\}$.

The successive sojourn times are denoted by $D_{l,i} = S_{l+1,i} - S_{l,i}, ...$

Then, we have the following result [see, e.g. Cinlar (1969), Cox (1970)]:

Proposition 1: The homogeneous Markov Chain assumption (i.e. Assumption A.2 a)) is satisfied if and only if:

i) The series of distinct successive states is Markov with transition matrix $\tilde{P}$ that contains the probabilities of each jump conditional on the previous state that is $\tilde{p}_{j,k} = p_{j,k}/(1 - p_{jj}),\ \forall j, k,\ j \neq k$ and $\tilde{p}_{jj} = 0,\ \forall j$.

ii) Conditional on the series of $\tilde{Y}_{l,i}, l = 0, 1, ...$, the sojourn times $D_{l,i}, l = 1, ..., n$ are independent, such that the distribution of $D_{l,i}$ is a geometric distribution with parameter p_{jj}, where $j = \tilde{Y}_{l,i}$.

Such an independence property of the sojourn times conditional on the occupied state characterizes a renewal process, and facilitates simulations as well as the derivation of asymptotic properties of the estimators. The homogeneous Markov chain is the benchmark model that can be extended in various aspects.

ii) **Time Dependent Chain**

Assumption A.2 b): $P_{i,t} = P_t$, independent of i and dependent on t.

This model allows us for taking into account the calendar time effects, such as the shortages of medical staff and hospital beds, the advances of knowledge about the disease or the fact that individuals get diagnosed earlier that helps them recover.

ii) **Chain with Duration Dependence**

Assumption A.2 c): $P_{i,t} = P_{t-S_{l_t,i}}$, where $S_{l_t,i}$ is the date of the most recent jump.

This transition depends on the time already spent in the current state.

Under Assumption A.2 c) the process maintains some properties of a renewal process, although the sojourn time distributions are no longer geometric distributions and $\tilde{Y}_{l,i}, l = 1, ..., n$ is no longer a Markov chain.

iv) **Chain with Cohort Effect**

Assumption A.2 d): $P_{i,t} = P_{S_0,i}$

The transition depends essentially on the diagnosis time. If we consider a given cohort (or generation) of individuals $\mathcal{P}_{s_0} = \{i : S_{0,i} = s_0\}$ composed of individuals diagnosed on the same date s_0, we have:

$P_{i,t} = P_{s_0}$, for $i \in \mathcal{P}_{s_0}$

Therefore, for each cohort, we have a homogeneous Markov chain, but the chain is no longer homogeneous if all cohorts are confounded due to some aggregation bias.

v) **Further extensions**

It is possible to consider more complex specifications by introducing jointly time dependence, duration dependence, a cohort effect and/or also the effects of individual characteristics such as the gender, age, co-morbidity, etc.

3 Descriptive Empirical Analysis of Transitions

This Section presents time independent summary statistics that can be easily computed by averaging individual histories over time and individual. The statistics discussed in this Section are transition frequencies and densities of sojourn times in a given state. These statistics are interpreted under the assumption of a homogeneous Markov chain. The time dependence is discussed at the end of this Section.

3.1 Transition Matrix

Below we provide the transition matrix P estimated from the entire sample.

Table 1. Estimated Frequency Matrix (%)

	D	ER	H	ICU	V	T	R	DE
D	96.04	0.03287	0.1237	0.02031	0.00481	0.02191	3.529	0.2271
ER	68.25	27.51	3.175	0	0	0	0.5291	0.5291
H	14.49	0.01337	81.62	0.4678	0.08019	0.4143	0.0401	2.874
ICU	9.242	0	5.0	81.21	0.303	1.591	0.07576	2.576
V	11.48	0	1.481	1.111	82.96	0.3704	0	2.593
T	4.813	0	0.4627	2.175	0.3702	88.99	0.09255	3.1
R	0	0	0	0	0	0	1	0
DE	0	0	0	0	0	0	0	1

The matrix can be interpreted as follows. Regardless of the individual characteristics the probability that an individual who is currently hospitalized is admitted to the ICU is

0.46% and that he/she dies is 2.87%, regardless of how long the patient has been in the hospital.

The six first lines of this transition matrix represent a "production process" of the medical care system, with inputs of infected individuals, and outputs being the recovered and deceased individuals. Along with the information on the time spent in each state, this is a basic tool for the budgeting of medical process, planning and scenario analysis [see, e.g. Alvarez et al. (2020), DiDomenico et al. (2020), Aguirregabiria et al. (2020), Atkeson (2020)].

3.2 Distributions of Sojourn Times

Let us now examine the durations of various states of medical treatment conditional on transitioning to another state. The analysis concerns the entire sampling period of 104 days.

We provide the sample distributions of sojourn times in the six states of interest. In order to (partly) eliminate the right censoring/truncation bias, the analysis concerns only complete sojourn times. For example, the distribution of a sojourn time in ICU conditional on exit to Intubation is estimated from individuals who have accomplished that transition.

Thus, the adjustment is carried out by examining the sojourn times conditional on the exit state. It simplifies the estimation at the expense of disregarding incomplete sojourn times and the decreasing numbers of individuals left at the end of the sample. This last effect is also due to delays in reporting.

The Figures 5-7 below show the distributions of sojourn times in H, ICU and T, conditional on the exit states. Additional summary statistics of these distributions (mean, variance, quantiles) are provided in Appendix A.3. These empirical distributions are evaluated from samples of different sizes, such as N=215 for Hospitalization before death, or N=6 for Hospitalization before Ventilation [see, Appendix A.3.]. In Figures 5-7, we only display the distributions evaluated from a sufficiently large number of observations.

The histograms show a decreasing pattern, except for the state Intubation illustrated in three panels of Figure 7. The exit state is unknown at the entry time into Intubation. However, it can be considered as a measure of severity of the disease with an effect of

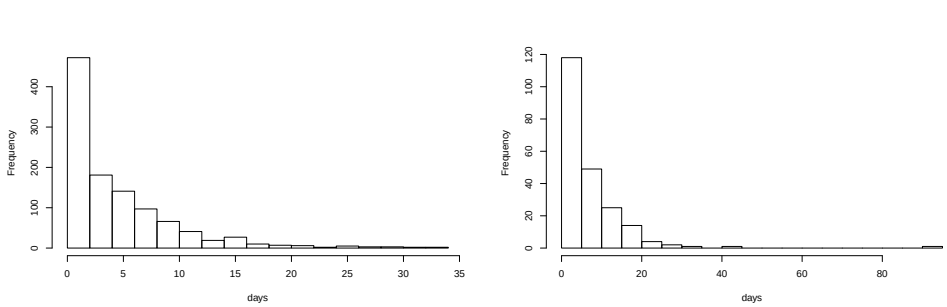


Fig 5. Durations of State Hospitalization Prior to Transition

the length of Intubation required. Under the assumption of perfect foresight of the future severity, we distinguish 3 severity levels for the intubated individuals: $v_1 = D, v_2 = ICU, v_3 = DE$.

These three distributions in Figure 7 feature fat right tails with 32 days of maximum Intubation. Depending on the severity level, 25% of individuals spend more than 8.5 days in level v_1, 16.5 days in level v_2 and 14.5 days in level v_3 [see Appendix A.3, 6]. We observe a similar pattern in the mean duration with 6.6 days in level v_1, 11.8 days in level v_2 and 9.1 days in level v_3. The decrease of the number of days between levels v_2 and v_3 reveals that a higher effort in medical care increases the probability of survival.

These sojourn time distributions are valid state by state and have to be used with caution for deriving conclusions on joint transitions. It is unlikely that an individual transitioning through states of H, ICU and T has an expected time of medical care equal to the sum of average times in H, ICU and T. There may exist a complicated negative or positive dependence between the sojourn times in the successive states. A patient with a severe condition may stay for a short time in ICU and a long time in T. Such complicated dependencies are better captured by means of a duration dependent transitions model [see, Section 4], then by an analysis of the joint distribution of sojourn times. For example, the correlation between the duration in H before exit to ICU and the duration in ICU before exit to Intubation is -0.128, the correlation between duration in ICU before exit to Intubation and duration in Intubation before Death is -0.233. These negative correlations confirm the discussion given above.

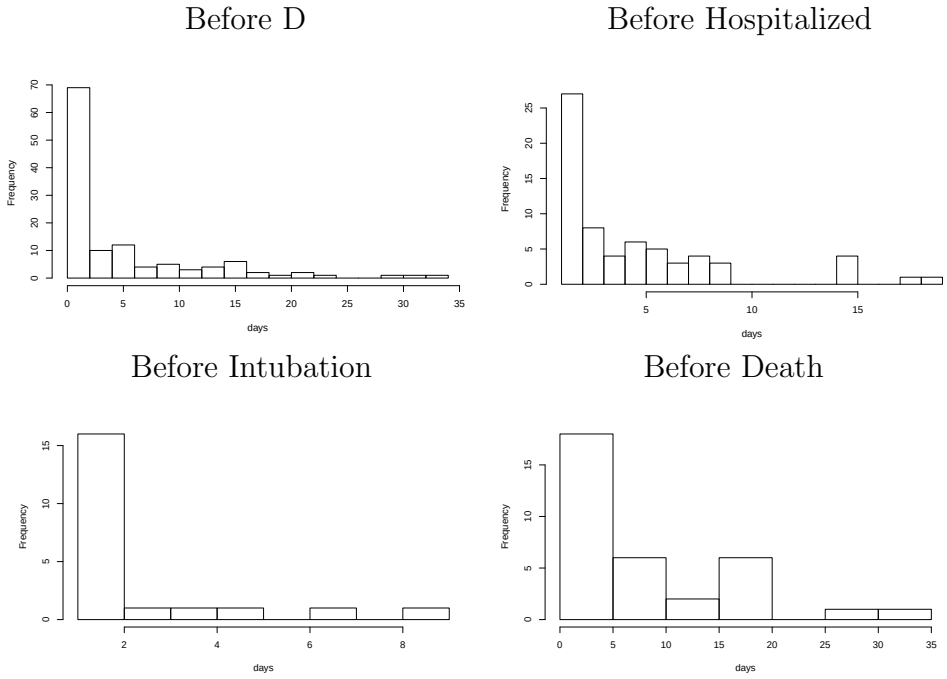


Fig 6. Durations of State ICU Prior to Transition

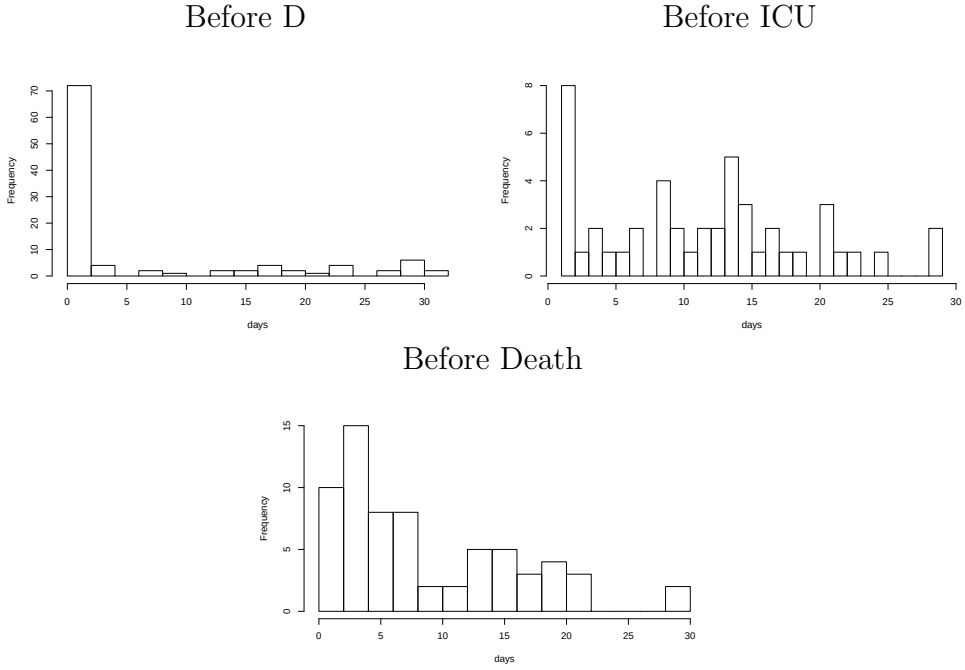


Fig 7. Durations of State Intubation Prior to Transition

3.3 Homogeneous Markov Chain Model

The summary statistics given in Sections 3.1 and 3.2 can be interpreted as evidence in favour or against the homogeneous Markov chain assumption A.2 a). In particular,

the empirical transition matrix given in Table 1 is the maximum likelihood estimator of matrix P in a homogeneous Markov Chain. Under A.2 a), the medical care process remains stationary and all non-stationarities in the counts of deceased and recovered are induced by the non-stationarity in the counts of diagnosed and their increase during the early phase of epidemic. This implies that the medical staff does not change their practices regarding the patient's treatment during the epidemic.

There is evidence in support of the homogeneous Markov chain assumption, such as the decreasing patterns of some duration densities, which resemble geometric densities and the fact that some of the densities do not depend on the state of exit. As well, the mean sojourn time computed directly from the duration data is often close to the mean sojourn time computed from the elements of the transition matrix in Table 1.

Let us, for example, consider the state Intubation with p_{jj} =0.89 (see, Table 1). Under Assumption A.2 a), the sojourn time follows a geometric distribution with mean $1/(1\text{-}p_{11})$ = 9.0 days, which can be compared with the expected values reported in Appendix A.3, 6.

On the contrary, the homogeneous Markov chain assumption seems inadequate for some states. Moreover, the sojourn times of the same individual in various states may be correlated, as mentioned above. Therefore, the assumption of homogeneous Markov Chain does not seem to hold, although it conveniently provides a simple framework for computing summary statistics that can serve as benchmarks for comparison [see, Section 4.3.1]. The alternative assumptions to replace A.2 a) are those of time dependence and duration dependence, for example.

Let us first examine if the medical care process is non-stationary. This can be done graphically by comparing the daily estimated transition matrices $\hat{P}_t$. By definition, the estimated transition probabilities $\hat{P}_{jk}$ in Table 1 are weighted averages:

$$\hat{P}_{jk} = \sum_{t=t_0}^{T-1} w_{jt} \hat{P}_{jk,t},$$

with weights that depend on the conditioning state and are proportional to he number of individuals in that state on day t.

Figure 8 shows the trajectories of daily transition probabilities $\hat{P}_{j,k,t}$ for selected states over the sampling period. The series have to be considered with caution as the number

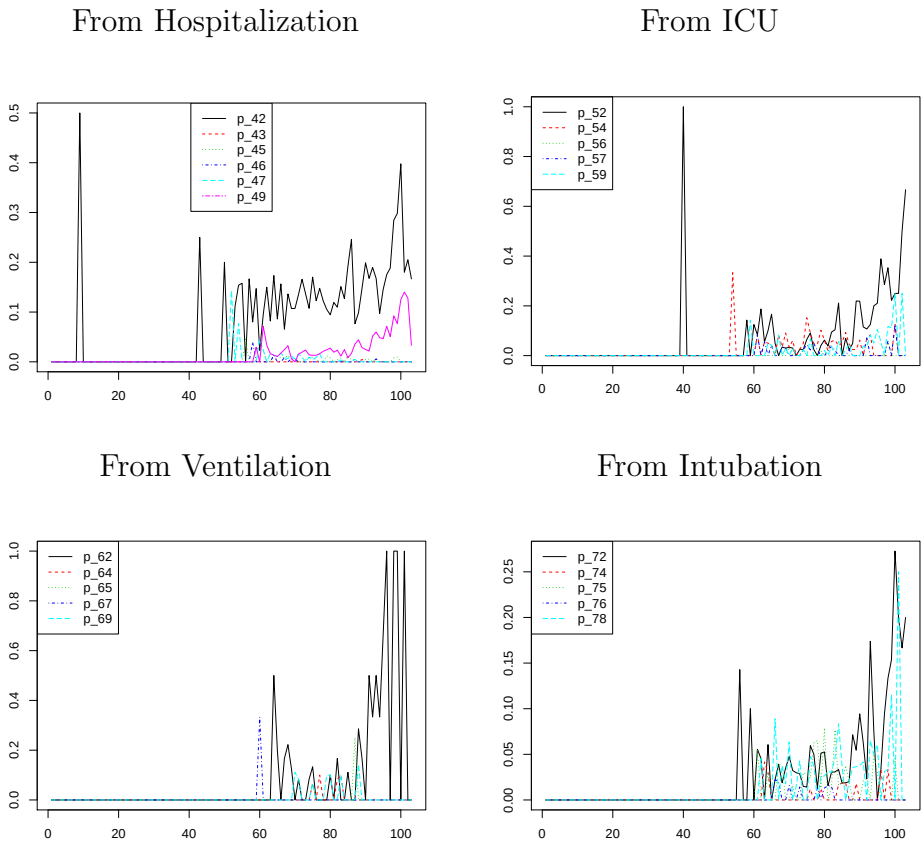


Figure 8 Transitions from States over 104 days

of individuals in a given state at a given date can be small. Therefore, we focus on the transitions from H and T to D (black lines), and from H and T to DE (pink and blue lines, respectively) computed from larger sub-samples. These series display upward trends, while the transitions from H to H and T to T (not reported) have downward trends by the unit mass restriction. Thus, the time spent in these states seems to decline. That can be due to a change in the treatment process, or to a left censoring effect. Indeed, the time of the diagnosis is not equal to the time of infection. If more tests were performed over time, infectious individual could be detected earlier and then it would be easier to cure them of COVID-19. The distinction between these alternative explanations is out of the scope of this paper.

Another remark concerns the evolution of the number of deaths. It is increasing smoothly from state H, which suggests it is being controlled. On the contrary, the evolu-

tion of deaths from state T is more erratic, which suggests there might be a pressure on the number of beds available.

4 Model with Duration Dependence

This Section extends the benchmark model by introducing duration dependence. First, we describe the specification of the transition matrix. Next, we provide the estimation results. These results are used for adjusting the count summary statistics for right truncation bias and for prediction making.

4.1 The Model

The (8×8) transition probability matrix P_t hat components $p_{j,k,i,t}$. The rows sum up to 1. Each row represents the probabilities of exit from a state j to another state k . The transition probabilities are specified as (conditional) multinomial logit models [see, McFadden (1984)]. The transition probabilities can depend on the time already spent in a state $Dur_{it}(D)$ They depend on the state and vary with both individual and time.

For example, let us illustrate the transitions $p_{j,k,i,t}$ from state j of D for individual i at time t:

$p_{1,1,i.t} = (\text{D} \rightarrow \text{D}) \propto 1$
$p_{1,2,i.t} = (\text{D} \rightarrow \text{ER}) \propto exp(\beta_{1,1} + \beta_{1,2} Dur_{it}(D))$
$p_{1,3,i,t} = (\text{D} \rightarrow \text{H}) \propto exp(\beta_{1,3} + \beta_{1,4} Dur_{it}(D))$
$p_{1,4,i,t} = (\text{D} \rightarrow \text{ICU}) \propto exp(\beta_{1,5} + \beta_{1,6} Dur_{it}(D))$
$p_{1,5,i,t} = (\text{D} \rightarrow \text{V}) \propto exp(\beta_{1,7} + \beta_{1,8} Dur_{it}(D))$
$p_{1,6,i,t} = (\text{D} \rightarrow \text{T}) \propto exp(\beta_{1,9} + \beta_{1,10} Dur_{it}(D))$
$p_{1,7,i,t} = (\text{D} \rightarrow \text{R}) \propto exp(\beta_{1,11} + \beta_{1,12} Dur_{it}(D))$
$p_{1,8,i,t} = (\text{D} \rightarrow \text{DE}) \propto exp(\beta_{1,13})$

where $Dur_{i,t}(D)$ denotes the duration (time) already spent in the current state D and the symbol $\propto$ denotes "proportional to", with the proportionality coefficient such that the above transition probabilities sum up to 1.

We mentioned earlier that state D comprises individuals of two types: D^1 prior to a potential medical care and D^2 -returned from medical care. Therefore the transition to a medical care concerns individuals in D^1 and depends on time spent in state $D = D^1$.

4.2 Estimation

The model is estimated by the Maximum Likelihood. The results of the estimation of the model are presented in Table 2, This Table contains the estimated coefficients for each row of the transition matrix. In each panel, the first coefficient is the intercept and the second one is the duration sensitivity.

The state "Recovery" is only observed at the end of the sample, and many individuals remain domiciled in the data while they could have recovered. Therefore the estimation of the transition from "D" to "Recovery", using the estimated parameters, is necessarily underestimated. Nevertheless, the dataset includes reliable information on the transition to/from the medical states and the state "Death". Thus, the estimates can be used to estimate accurately the dynamic of the probabilities and counts in these states.

Table 2: Estimated Coefficients

	TRANSITION FROM D												
States	ER		H		ICU		V		T		R		DE
Parameters	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{1,6}$	$\beta_{1,7}$	$\beta_{1,8}$	$\beta_{1,9}$	$\beta_{1,10}$	$\beta_{1,11}$	$\beta_{1,12}$	$\beta_{1,13}$
Estimates	-8.0924	0.0040	-5.0664	-0.0665	-6.9397	-0.0635	-9.1849	-0.0262	-6.4233	-0.0861	-2.7289	-0.0217	-6.0468
	TRANSITION FROM ER												
States	D	H	R	DE									
Parameters	$\beta_{2,1}$	$\beta_{2,2}$	$\beta_{2,3}$	$\beta_{2,4}$									
Estimates	0.9086	-2.1595	-3.9512	-3.9512									
	TRANSITION FROM H												
States	D		ER		ICU		V		T		R		DE
Parameters	$\beta_{3,1}$	$\beta_{3,2}$	$\beta_{3,3}$	$\beta_{3,4}$	$\beta_{3,5}$	$\beta_{3,6}$	$\beta_{3,7}$	$\beta_{3,8}$	$\beta_{3,9}$	$\beta_{3,10}$	$\beta_{3,11}$	$\beta_{3,12}$	$\beta_{3,13}$
Estimates	0.2103	-0.2535	-1.6251	-3.7254	-3.3540	-0.2278	-3.4636	-0.7252	-3.0971	-0.3065	-8.1585	0.0299	-3.3466
	TRANSITION FROM ICU												
States	D		H		V		T		R		DE		
Parameters	$\beta_{4,1}$	$\beta_{4,2}$	$\beta_{4,3}$	$\beta_{4,4}$	$\beta_{4,5}$	$\beta_{4,6}$	$\beta_{4,7}$	$\beta_{4,8}$	$\beta_{4,9}$	$\beta_{4,10}$	$\beta_{4,11}$		
Estimates	-0.2942	-0.2116	-0.7782	-0.2372	-4.2901	-0.1215	-1.6329	-0.3035	-7.9989	0.0593	-3.4509		
	TRANSITION FROM V												
States	D		H		ICU		T		R		DE		
Parameters	$\beta_{5,1}$	$\beta_{5,2}$	$\beta_{5,3}$	$\beta_{5,4}$	$\beta_{5,5}$	$\beta_{5,6}$	$\beta_{5,7}$	$\beta_{5,8}$	$\beta_{5,9}$	$\beta_{5,10}$	$\beta_{5,11}$		
Estimates	0.7679	-0.2843	-2.2051	-0.1448	-1.8467	-0.2322	7.5903	-9.7537	-53.9343	-66.3286	-3.4657		
	TRANSITION FROM T												
States	D		H		ICU		V		R		DE		
Parameters	$\beta_{6,1}$	$\beta_{6,2}$	$\beta_{6,3}$	$\beta_{6,4}$	$\beta_{6,5}$	$\beta_{6,6}$	$\beta_{6,7}$	$\beta_{6,8}$	$\beta_{6,9}$	$\beta_{6,10}$	$\beta_{6,11}$		
Estimates	-0.3381	-0.2170	-3.3238	-0.1377	-2.1517	-0.1031	-3.8112	-0.1127	-7.5641	0.0351	-3.3569		

The duration dependence is introduced for each state except the state of emergency room ER. Indeed, a majority of individuals stay in ER for one day only. Table 2 does not report the results concerning probabilities of remaining in a given state, which are assumed to be proportional to 1, for identification. The values of parameters of a multinomial model do not have a direct interpretation. Instead the ratios of transition probabilities are interpretable (in the 2-state case, the odd-ratios are interpretable).

Let us consider the state intubation T. After one day spent in this state, the ratio of the transition to D (i.e. the way to recovery) and the transition to death DE are:

$$\exp(\beta_{6,1} + \beta_{6,2} - \beta_{6,11}).$$

After 10 days, this ratio is:

$$\exp(\beta_{6,1} + 10\beta_{6,2} - \beta_{6,11}).$$

Therefore, the ratio of these two ratios 10 days/1 day is:

$$\exp(9\beta_{6,2}) = \exp(-0.22 \times 9) < 1.$$

Thus, when the time spent in Intubation T increases, the chance to exit on the way to recovery diminishes, as compared to the earlier days.

We illustrate these effects conditional on the state sojourn time (duration), which we set to vary from 1 to 15 days in Figures 15-19 in Appendix 4.

Figure 15 depicts the dynamic of the transition probabilities from the state "D" in terms of its duration. This figure contains eight panels. Each panel shows the probability of transition to the states: D, ER, H, ICU, V, T, R, and DE. The probability that someone in state D stays in state D remains above 0.93 and increases with the duration of the isolation. Indeed, isolated individuals mostly have mild symptoms and are more likely to stay isolated until their recovery. The probability of recovering varies between 0.060 after 1 day and 0.045 after 15 days. However, we expect this probability to be higher due to the lack of observability of the date of recovery. The transition probabilities to the other states are negligible.

Figure 16 displays the transitions from hospitalization. Individuals with a day of hospitalization likely have mild symptoms and have a probability of 0.46 to be isolated. This probability decreases as the duration increases while the risk to remain hospitalized increases. Similarly, Figures 17-19 show the transition probability from the states "ICU", "Ventilation" and "Intubation".

The probability of remaining in a state increases quickly with the time spent in T. The probability of death seems to increase with time spent in all types of medical care,

at rates depending on the intensity of care. The rate of increase is the highest for V, followed by that from ICU, followed by a slightly lower rate from T.

In general, the probabilities of transition to death after long durations are close to the transition probabilities in Table 1, computed under the assumption of homogeneous Markov chain.

The results concerning the probability to recover are affected by the lack of the date of recovery and need to be interpreted with caution.

4.3 Fitted Values and Predictions

Let us now explain how the estimated transition model can be used to compute the fitted values of counts over the observation period and the forecasts. First, we consider the computation under the assumption of homogeneous Markov chain and next under the assumption of duration dependence. We focus on the counts of hospitalized, ICU and Intubation.

4.3.1 **Homogeneous Markov Chain**

Recall that state D includes two types of histories: D^1 are individuals who are isolated at the detection date and D^2 are individuals who are on their way to recovery after a transition through H, ICU, V and T. For subsequent computation, we need to separate these two types of observations and focus on D^1. The first row of the transition matrix in Table 1 is replaced by:

Table 3. Modified Row 1 of Frequency Matrix (%)

	D^1	ER	H	ICU	V	T	R	DE
D^1	93	0.0	1	0.2	1	0.2	3.6	1

The modified transition matrix is denoted by $\tilde{P}$. After being diagnosed, an individual is assigned to a state. The time-independent (constant) probabilities of assignments estimated from the entire sample are:

Table 4. Probabilities of Assigment (%)

	D	ER	H	ICU	V	T	R	DE
v	93	1.0	5.0	1.0	0	0	0	0

We have a causal scheme with 3 episodes:

1. the detection episode that ends at the case reported date.

2. the assignment episode, with the selection based on v,

3. the medical care process summarized by matrix $\tilde{P}$.

At time T=May 05, we get for each individual either a recovery, or a death or a censored duration in the last state, such as one of the medical care states.

The fitted values, i.e. the expected counts, are computed from the sequence of inputs X_t, which are the counts of new diagnosed on day t (see Figure 2). Let N_t denote the row vector of length 8, with elements equal to the count of individuals in each state on day t. The fitted counts are given by:

$$EN_t = \sum_{\tau=1}^{t} (X_\tau v' \tilde{P}^{t-\tau}), \quad t = 1, ..., 104, \tag{4.1}$$

Thus, we evaluate the future expected counts for each cohort (generation) τ and sum over the past cohorts. Figure 9 below provides the fitted counts of H, ICU and T. They are close to the patterns shown in Figure 3 up to day 80, and are significantly different afterwards. Indeed, the fitted values provide an adjustment for the truncation bias revealed in Figure 3. It is therefore important to use the model-based figures instead of data-based figures in the presence of truncation and reporting lag problems, to avoid misleading conclusions, especially concerning the peak of the epidemic. Figure 9 shows the peak of Hospitalization that seems about to appear at the beginning of May for state H (from recent data, we know that there was a peak in the first wave, followed by a peak in the second wave). The peak in Intubation is not visible from the fitted counts due to the significant time spent in intubation T.

4.3.2 **Model with Duration Dependence**

A similar analysis can be performed by using the model with duration dependence discussed in Sections 4.1-4.2. Then equation (4.1) becomes:

$$EN_t = \sum_{\tau=1}^{t} (X_\tau v' E[\tilde{P}_{\tau+1}, \tilde{P}_{\tau+2} \cdots \tilde{P}_t)]), \quad t = 1, ..., 104. \tag{4.2}$$

It accounts for the fact that the matrices depend on time through the durations spent in each state. As the individual durations Dur_{it} are random, the product of matrices has to be integrated out, which can be easily done by simulations.

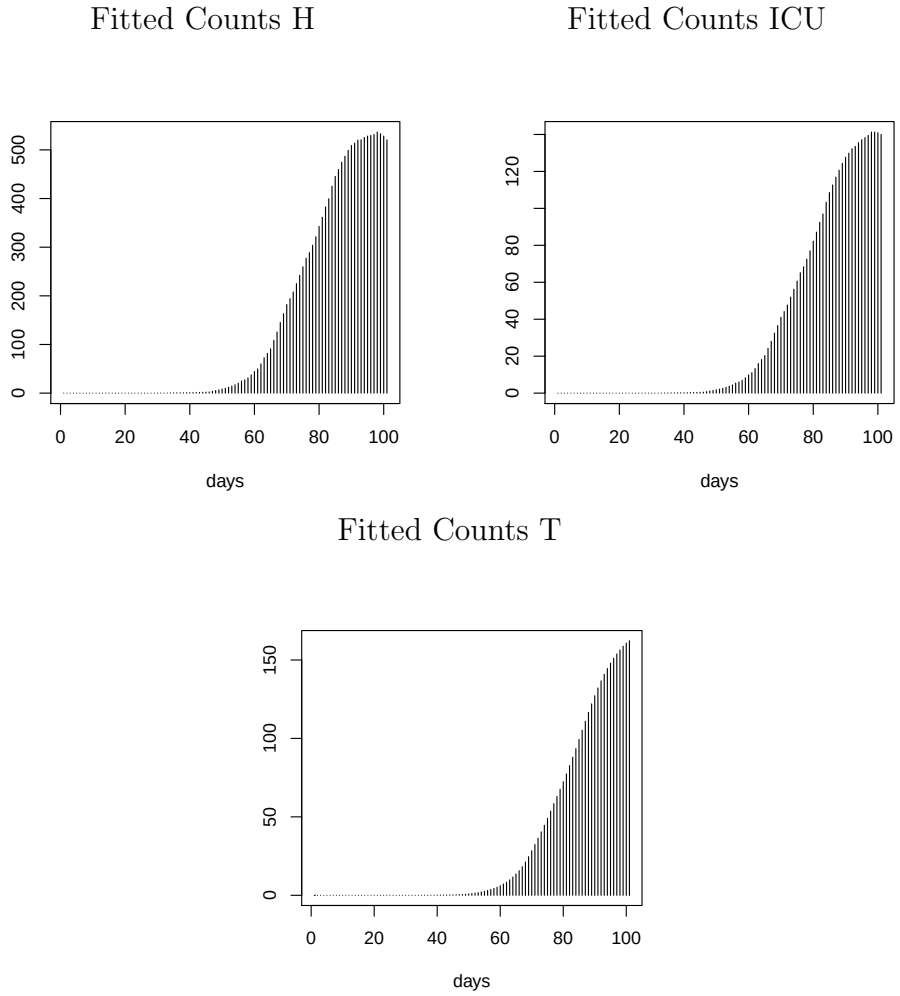


Figure 9. Fitted Counts of Medical Care over 104 days

Figures 10 and 11 show the estimated and observed counts of patients under medical care and the cumulative counts of deaths, respectively. Figure 10 illustrates the good fit of the transition model to the data and the satisfactory predictive power of the model. The model predicts, as expected from Figure 3, higher medical state counts for the remaining dates. Figure 11 plots the estimated and the observed number of cumulative deaths. The figure shows that the total death is accurately predicted. The figures confirm that the model provides an adjustment for truncation and reporting lag. As mentioned earlier the data suffer from a truncation bias and a lag in reporting that start around day 80 of the observed data and generate decreasing patterns of daily counts, found ex-post

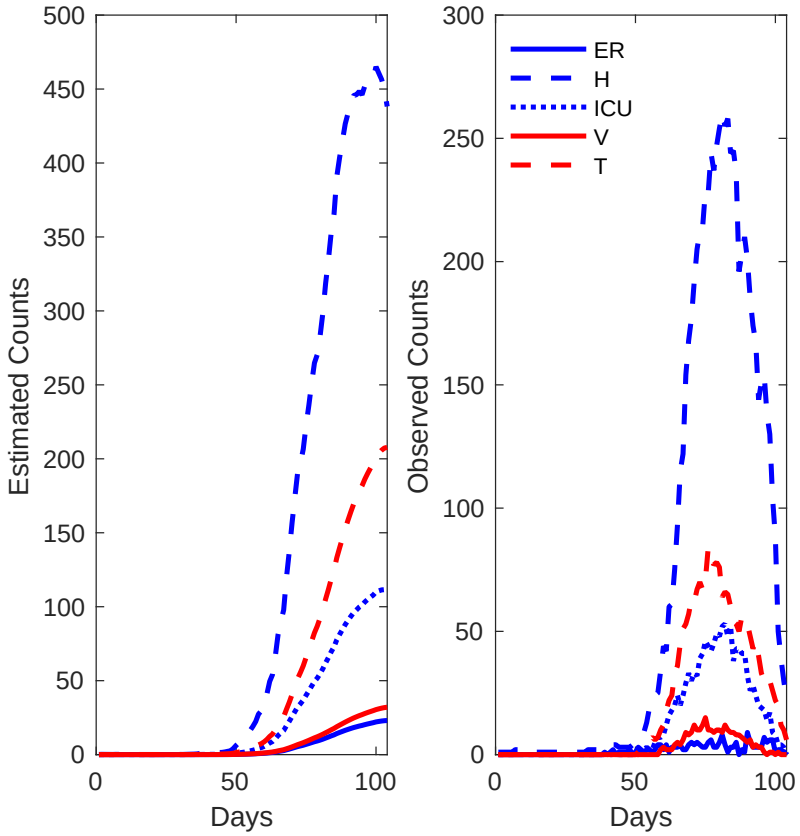


Figure 10: Estimated and Observed Counts of Medical Care

incompatible with the publicly available counts provided later by the PHO. The model is shown to "adjust" for these biases.

4.3.1 **Predictions**

A similar approach can be applied to predict the future counts and the numbers of beds required for the different types of medical care.

Due to the time of medical treatment, we recommend predicting a term structure of counts, up to say 20 days rather than computing one day ahead forecasts only. This long-run prediction requires additional forecasting of future inputs $X_{T+1}, X_{T+1}, ..., X_{T+H}$. This is only possible if the model is completed by a dynamic model for the X_t's, from a SI component of the SIRD-type of model, obtained by logit adjustments for the series of cumulated counts of diagnosed individuals. This is clearly out of the scope of the present

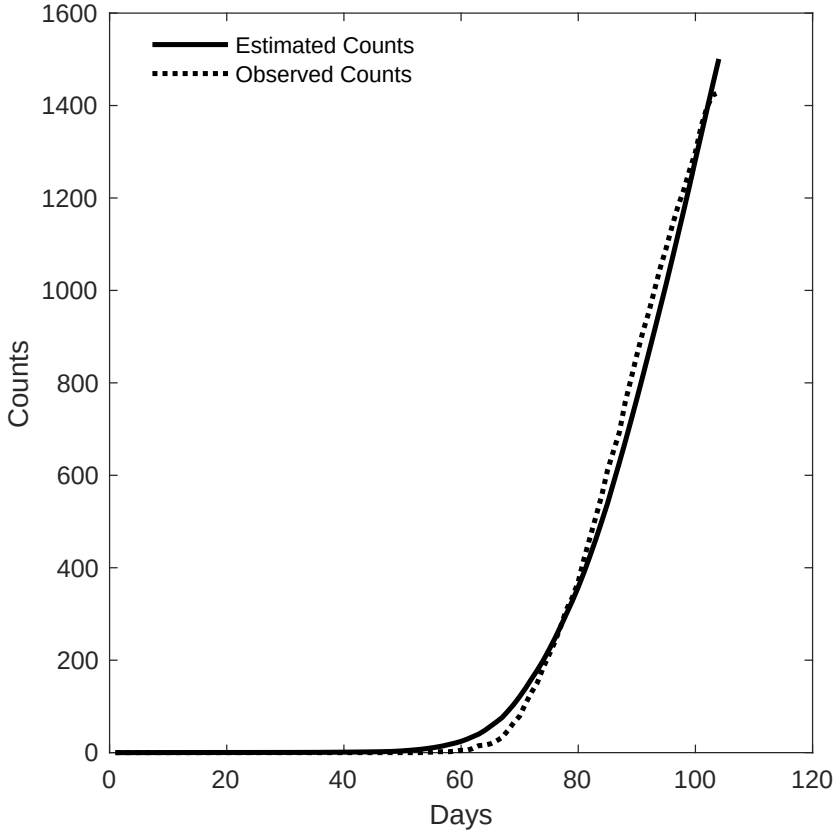


Figure 11: Estimated and Observed Counts of Deaths

paper.

5 Concluding Remarks

This paper examined the process of medical treatment during the early phase of COVID-19 epidemic in Ontario. We investigated the medical care dynamics and its adjustment over time. We considered a benchmark transition model based on the assumption of homogeneous Markov chain and studied its extension that accounts for time dependence and duration dependence. The advantage of the transition model is the adjustment for truncation that helps eliminate misleading results based on a naive interpretation of the series of counts.

We have disregarded the effect of the detection process of infected individuals. Indeed, the process may induce both left truncation and left censoring. The left truncation arises as in the case of COVID-19 the analysis is complicated further by the large number of asymptomatic individuals who are difficult to detect [see e.g. Gourieroux, Jasiak (2020) for an approach to recovering the unobserved counts of undetected infectious individuals]. The left censoring arises, as the detection date does not coincide with the infection date, and the difference between these dates may depend on the detection efforts and the numbers of tests performed.

The transition model-based analysis can be further extended to accommodate the individual characteristics and spatial dependence between the infections in Ontario.

REFERENCES

Alvarez, F., Argente, D. and F. Lippi (2020): A Simple Planning Problem for COVID-19 Lockdown, LUISS and Einaudi Institute for Economics and Finance DP.

Aguirregabiria, V., Gu, J., Luo, Y. and P. Mira (2020): A Dynamic Structural Model of Virus Diffusion and Network Production: A First Report. Working Paper 665, University of Toronto.

Atkeson, A. (2020): What Will Be The Economic Impact of Covid-19 In The US? Rough Estimates of Disease Scenarios, NBER Working Paper 26867

Blanghaps, N. Nov, Y. and G. Weiss (2013): Sojourn Times Estimation in and M/G/∞ Queue with Partial Information, Journal of Applied Probability, 27, 314-324.

Berkson, J. (1944): Application of the Logistic Function to Bio-Assay, JASA, 339-357.

Cinlar, E. (1969): Markov Renewal Theory.

Cox, D. (1970): Renewal Theory, London

DiDomenico, L., Pullano, G., Coletti, P., Hens, N. and V. Colizza (2020):Expected Impact of School Closure and Telework to Mitigate COVID-19 Epidemic in France, www.epicx-lab.com/covid-19.html]

Gourieroux, C., and J. Jasiak (2007): Econometrics of Individual Risks: Credit, Insurance and Marketing, Princeton University Press.

Gourieroux, C., and J. Jasiak (2020): Time Varying Markov Processes with Partially Observed Aggregate Data: An Application to Coronavirus, Arxiv 2005.04500.

Grasselli, G., Pesenti A., and Cecconi, M. (2020): Critical Care Utilization for the Covid-

19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response, Journal of the American Medical Association, forthcoming.

Hens, N, Shkedy, Z., Aerts, M., Faes, C.,Van Damme, P. and P. Beutels (2012) Modeling Infectious Disease Parameters Based on Serological and Social Contact Data, Springer, New York.

Kermack, W., and A., McKendrick (1927): A Contribution to the Mathematical Theory of Epidemics,Proceedings of the Royal Statistical Society, A, 115,700-721.

Lapidus, N., Zhou, X., Carrat, F. Riou, B. , Zhao, Y., and G. Hejblum (2020): Biased and Unbiased Estimation of the Average Length of Stay in Intensive Care Unit in the Covid-19 Pandemic, medRxiv.

McFadden, D. (1984): Econometric Analysis of Qualitative Response Models,in Handbook of Econometrics,Vol 2,1395-1457, Elsevier

Murray, C. (2020): Forecasting COVID-19 Impact on Hospital Bed-Days, ICU-Days, Ventilator-Days by US State in the Next 4 Months, Medrxiv, March 30.

Office of the Premier, News, April 16, 2020, 1:00 pm: Ontario Significantly Expands Hospital Capacity to Prepare for Any COVID-19 Outbreak Scenario.

Public Health Ontario (PHO) 2020a: Enhanced Epidemiological Summary: COVID-19 and Severe Outcomes in Ontario, April 22 https://www.publichealthontario.ca/-/media/documents/ncov/epi/covid-19-severe-outcomes-ontario-epi-summary.pdf?la=en

Public Health Ontario (PHO) 2020b: Epidemiological Summary: COVID-19 in Ontario: January 15, 2020 to May 27, 2020 https://www.publichealthontario.ca/-/media/documents/ncov/epi/2020/covid-19-daily-epi-summary-report.pdf?la=en

Public Health Ontario (PHO) 2020c: Enhanced Epidemiological Summary: COVID-19

in Ontario: A Focus on Diversity, May 14, 2020 https://www.publichealthontario.ca/-/media/
documents/ncov/epi/2020/covid-19-daily-epi-summary-report.pdf?la=en

Schweer, S. and C. Wickelhaus (2015): Nonparametric Estimation of the Service Time Distribution in the Discrete Time GI/G/∞ Queue with Partial Information, Stochastic Processes and Their Applications, 125, 233-253.

Toda, A. (2020): Susceptible-Infected-Recovered (SIR) Dynamics of COVID-19 and Economic Impact, ArXiv:2003.11221v2

Vynnicky,.E., and R. White (eds)(2010): An Introduction to Infectious Disease Modelling, Oxford Univ Press

Yan, P., and G., Chowell (2019): Quantitative Methods for Investigating Infectious Disease Outbreaks, Springer.

Appendix A.1

IPHIS_REPORT.csv DATASET

This Appendix provides the codes of variables in the database.

DETECTION

These data describe the phase of detection and provide basic individual characteristics.

A: source - iPHIS (Integrated Public Health Information System) or CORES (Toronto Public Health Coronavirus Rapid Entry System)

B: client ID

C: case ID

D: client first name - coded

E: client second name - coded

F: client last name - coded

G: client DOB (Date of Birth)

H: outbreak number

I: classification description confirmed

J: case status description open or closed

K: disposition description pending or complete

L: case created date

M: case reported date - either same as L or before L

N: specimen date

O: accurate episode date - after N

P: symptom onset date - mostly = O

Q: diagnosing health unit area - Toronto, York, Peel Region, etc.

R: client address city - Mississauga, Thornhill, Toronto, etc.

S: client address - coded

T: client postal code - given

U: client province

V: age at time of illness

W: client gender

MEDICAL CARE

These data describe the types of medical care of each patient (referred to as client in the database).

X: hospitalized - yes, no, unknown

Y: hospitalization admit date

Z: hospital discharge date

AA: ICU - yes or no

AB: ICU start date

AC: ICU end date

AD: ER (Emergency Room) visit - yes or no

AE: ER visit start date

AF: ER visit end date

AG: intubation - yes or no

AH: intubation start date

AI: intubation end date

AJ: isolation - yes, no, unknown

AK: isolation start date - before hospitalization or ER

AL: isolation end date

AM: ventilator - yes, no, unknown

AN: ventilation start date

AO: ventilation end date

AP: case acquisition info - travel related, contact of confirmed case, neither, or info pending

AQ: outcome - recovered, pending, fatal

AR: resolved - yes, no, fatal

AS: client death date

AT: death cause - unknown, reportable disease contributed, but was not underlining cause/ was underlining cause

COMPLICATIONS FROM DISEASE:

The knowledge of complications is important to evaluate ex-post the cause of death.

AU: compx ards (deadly lung condition) - yes, no, not asked

AV: compx encephalitis (brain inflammation) - yes, no, not asked

AW: compx pneumonia - yes, no

AX: compx renal failure (kidney failure) - yes, no, not asked

AY: compx mening nuchal rigidity (inability to flex the neck from muscle spasm) - appears empty due to missing values

AZ: compx liver failure - yes, no but mostly empty

BA: compx heart failure - yes, no but mostly empty

BB: compx respiratory failure - yes, no but mostly empty

BC: compx palpitation arrhythmia - yes, no, not asked

BD: compx sepsis (body's response to infection damages organs) - yes, no, not asked

BE: compx other

BF: asymptomatic - yes, no, unknown

SYMPTOMS:

The symptoms provide insights on the severity of the disease at detection time.

BG: sym cough - yes, no

BH: sym fatigue - yes, no

BI: sym fever - yes, no

BJ: sym headache - yes, no

BK: sym shortness of breath - yes, no

BL: sym sore throat - yes, no

BM: sym other

PRE-EXISTING HEALTH CONDITIONS:

These health conditions include risk factors that can have an impact on the severity of infections, i.e. the co-morbidities. These include asthma, obesity, diabetes and weakness of immune system due to a cancer treatment.

BN: anemia or hemoglobinopathy - yes, no

BO: asthma - yes, no, unknown

BP: cancer - yes, no, unknown

BQ: cardiovascular conditions - yes, no, unknown

BR: UMC (Urine Microscopy Culture) - yes, no, unknown

BS: chronic liver disease - yes, no, unknown

BT: COPD (Chronic Obstructive Pulmonary Disease) - yes, no, unknown

BU: diabetes - yes, no, unknown

BV: immune compromised - yes, no, unknown

BW: neurologic disorder - yes, no, unknown

BX: obesity - yes, no, unknown

BY: postpartum - yes, no, unknown

BZ: pregnant - yes, no

CA: renal conditions - yes, no, unknown

CB: tuberculosis - yes, no, unknown

CC: other illnesses - yes, no, unknown

SOURCE OF CONTAMINATION:

This is basic information for tracking the infection

CD: close contact - yes, no, unknown

CE: travel RF - yes, no

CF: OCC HCW (occupation of healthcare worker) - yes, no

CG: CCF LTCHRES (Complaint, Critical Incident and Follow-Up at long term care home residence) - yes, no, unknown

CH: alcohol abuse- yes, no, unknown

CI: close contact with animals - yes, no, unknown

CJ: homelessness - yes, no, unknown

CK: H (human? home? close contact) - yes, no, unknown

CL: IDU (injection drug user) - yes, no, unknown

CM: nosocomial (disease caught in hospital) - yes, no, unknown

OCCUPATION:

The information about the professional occupation (occ) is summarized by the following variables.

CN: HCW (health care worker) - yes, no

CO: occ animal oraph (Ontario Animal Protection Hotline worker) - yes, no

CP: occ DR (doctor) - yes, no

CQ: occ farmworker - yes, no

CR: occ first responder - yes, no

CS: occ municipal worker - yes, no

CT: occ nurse - yes, no

CU: occ labworker - yes, no

CV: occ shelter homeless staff - yes, no

CW: occ vet (veterinary) - yes, no

ADDITIONAL INFORMATION

CX: smoker

CY: visited HCF (health care facility) - yes, no

CZ: placer requisition ID - coded

DA: test reported date

DB: ETL load date (May 06, 2020 is the date of dataset being uploaded)

Appendix A.2

Data Adjustments

We have implemented the following data adjustments prior to estimation:
1) States are determined by the availability of the beginning date or the final date in the states.

2) The Case Reported Dates are used as the date of detection. The first date of the reported detected case in our data is January 23rd, 2020. The last date is May 5th, 2020.

3) We constructed the history of individuals from their detection date to May 5th.

4) Some cases were hospitalized for other reasons many weeks before the outbreak and the detection dates. The relevant COVID hospitalization period is considered as the one after the COVID-19 detection date.

5) For the individual in row 13306, dates were reported with an error. We treat these dates as missing information.

6) The individual in row 4506 has a missing value for a death date. We set it to the final date.

7) If the first or the last date in a state is missing, we assume the individual spent one day in that state. For instance, most people in our data who went to the emergency room have only one date reported. This is consistent with the fact that people often return the same date from the emergency room.

8) The beginning and end dates in states were often inverted. We fixed that.

9) For individuals in rows 6425, 6596, and 14582, the start and end dates in the emergency room were 18 MARCH 2020 to 18 APRIL 2020, 20 MARCH 2020 to 20 APRIL 2020, and 14 MARCH 2020 to 14 APRIL 2020, respectively. This corresponds to a month in the emergency room and cannot be accurate. We assumed that the terminal period month should be MARCH and corrected the end date.

Appendix A.3

Sojourn Times - Descriptive Statistics

1. State D

Duration in D before ER.

There are N =123 patients with only 2 durations over the first 30 days.

	N	min	25%	median	75%	max	mean	var
total	123	1	2	3	8	40	5.813	37.940

Duration in D before Hospitalized

	N	min	25%	median	75%	max	mean	var
total	463	1	2	4	8	36	5.460	20.309

Duration in D before ICU

There are only 5 cases over the first 30 days. Individual 16100 spends one day to ICU and returns to state D.

	N	min	25%	median	75%	max	mean	var
total	77	1	2	3.5	6	32	4.644	19.432

Duration in D before Ventilation

There are 18 individuals who move from D to Ventilation. These transitions took place after day 40, i.e. about April 02.

	N	min	25%	median	75%	max	mean	var
total	18	1.0	2.0	4.0	6.5	26.0	5.611	40.722

Duration in D before Intubation

	N	min	25%	median	75%	max	mean	var
total	82	1	2	4	6	14	4.365	8.185

Duration in D before Recovered.

These individuals may have undergone medical treatments, returned to D and were reported as Recovered.

	N	min	25%	median	75%	max	mean	var
total	13207	1	17	24	33	103	25.147	109.524

Duration in D before Deceased

There are no transitions from D to Deceased over the first 62 days. Individual 14815 dies after 49 days in isolation without any medical treatment.

	N	min	25%	median	75%	max	mean	var
total	851	1	4	7	11	49	7.994	35.323

The durations prior to transitions from state 3 of De are given in Figure 5.

2. State ER

Duration in ER before return to D

	N	min	25%	median	75%	max	mean	var
total	137	1	1	1	1	25	1.781	12.157

Duration in ER before Hospitalization

The results are based on N=6 patients.

	N	min	25%	median	75%	max	mean	var
total	6	1	1	3	5	7	3.333	7.066

Individual 3498 has recovered on the last day of sampling period on day 104 after staying in ER. Individual 17382 died after 2 days in ER.

3. State Hospitalization

Duration in Hospitalization before return to D

	N	min	25%	median	75%	max	mean	var
total	1084	1	1	3	7	34	4.959	27.008

Duration in Hospitalization before ER

Individual 8654 made that transition after 1 day in hospital.

Duration in Hospitalization before ICU

These durations are recorded after day 57 of the sample, i.e. about March 20.

	N	min	25%	median	75%	max	mean	var
total	35	1.0	1.0	2.0	3.5	15.0	3.085	10.668

Duration in Hospitalization before Ventilation.

These transitions start on day 59 of the sampling period, about March 22.

	N	min	25%	median	75%	max	mean	var
total	6	1.00	1.25	2.50	3.00	3.00	2.166	0.966

Duration in Hospitalization before Intubation

These transitions start on day 53 , i.e. about March 16

	N	min	25%	median	75%	max	mean	var
total	31	1	1	2	3	12	2.774	6.780

Duration in Hospitalization before Recovered

There are only 3 transitions after hospitalization of 28, 9 and 39 days.

Durations of Hospitalization before Death

The are no transitions to state 9 from Hospital over the first 30 days.

	N	min	25%	median	75%	max	mean	var
total	215	1	3	5	10	92	7.525	71.568

4. State ICU

Duration in ICU before D

There are only 2 transitions before day 60 of March 23.

	N	min	25%	median	75%	max	mean	var
total	122	1.00	1.00	1.00	6.75	33.00	5.327	47.627

Duration in ICU before Hospitalization

There is only one transition before day 60.

	N	min	25%	median	75%	max	mean	var
total	66	1	2	3	6	19	4.863	18.673

Duration in ICU before Ventilation

There are only 4 durations between days 71 (April 3) and 99 (April 30) of 1,3 ,6 and 7 days.

Duration in ICU before Intubation

There are no transitions before day March 23 and only 4 durations between April 22 and the end of sample

	N	min	25%	median	75%	max	mean	var
total	21	1	1	1	2	9	2.285	4.914

Duration in ICU before Recovered

Individual 9614 makes a transition into R on day 104 (May 04) after 20 days in ICU.

Duration in ICU before Death

There are no transitions before day 60 (March 23).

	N	min	25%	median	75%	max	mean	var
total	34	1	2	5	14	35	4.863	18.673

5. State Ventilation

Duration in Ventilation before return to D

There are no durations before day March 23 (day 60).

	N	min	25%	median	75%	max	mean	var
total	31	1	1	1	2	26	4.774	57.580

Duration in Ventilation before return to Hospitalized

There are only 4 durations between days 76 and 84 of 2,8,1, and 25 days.

Duration in Ventilation before return to ICU

There are 3 durations between days 76 and 88 of length 1, 8 and 9.

Duration in Ventilation before Intubation

There is only one 1-day duration on day 61 (March 24).

Duration in Ventilation before Death

	N	min	25%	median	75%	max	mean	var
total	7	5.0	7.5	9.0	11.0	16.0	9.571	12.952

6. State Intubation

Duration in Intubation before return to D

There is only one transition before day 60 (March 23).

	N	min	25%	median	75%	max	mean	var
total	104	1.0	1.0	1.0	8.5	32.0	6.586	96.380

Duration in Intubation before return to Hospitalized

There are no transitions before day 60 (March 23).

	N	min	25%	median	75%	max	mean	var
total	10	1.00	7.50	10.50	11.75	17.00	9.9	22.988

Duration in Intubation before return to ICU

There are no transitions before day 60 (March 23).

	N	min	25%	median	75%	max	mean	var
total	47	1.0	5.5	12.0	16.5	29.0	11.80851	59.679

Duration in Intubation before Ventilation

All durations occurred between days 67 and 83.

	N	min	25%	median	75%	max	mean	var
total	8	1.0	1.0	10.0	18.5	29.0	11.25	.

Duration in Intubation before Recovered

There are 2 durations with transition on day 104 of 35 and 7 days.

Duration in Intubation before Death

	N	min	25%	median	75%	max	mean	var
total	67	1.0	3.0	7.0	14.5	30.0	9.149	53.886

Appendix A.4

Additional Figures

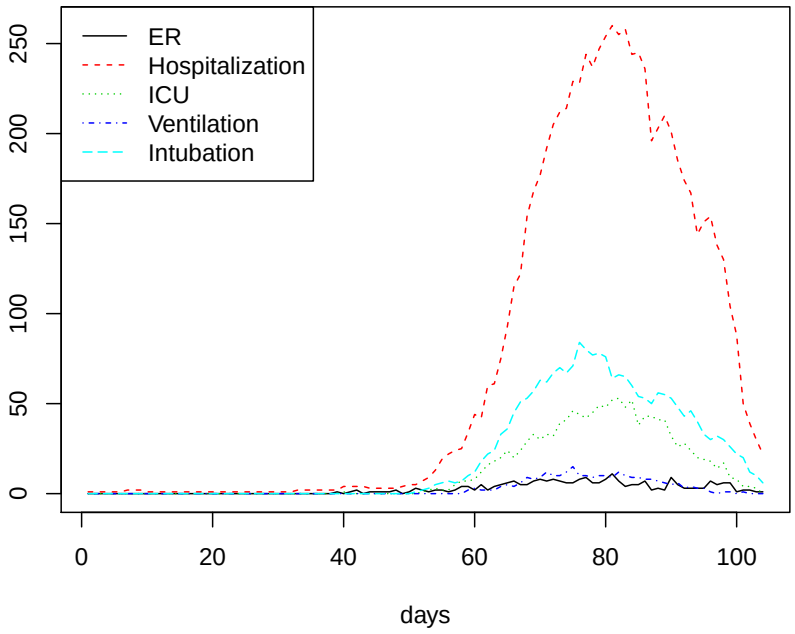


Figure 12: Counts of Medical Care States 3 to 7

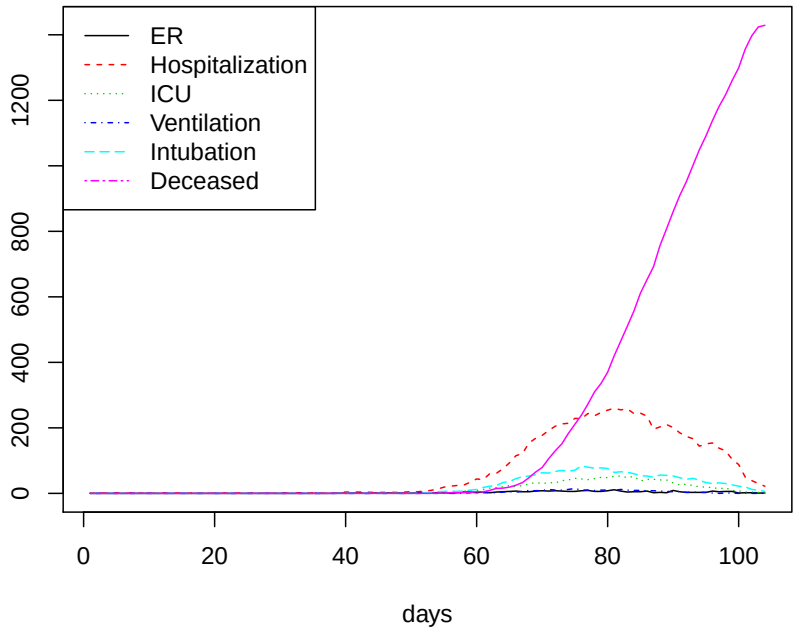
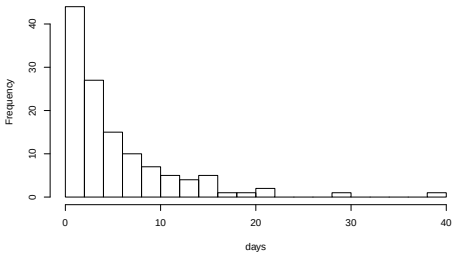
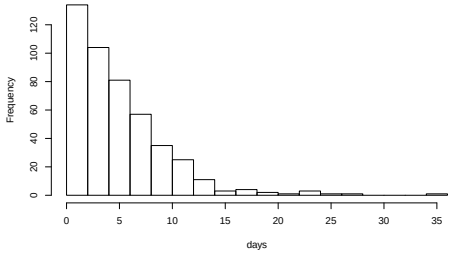


Figure 13: Counts of Medical Care States 3 to 7 and Cumulated Deceased

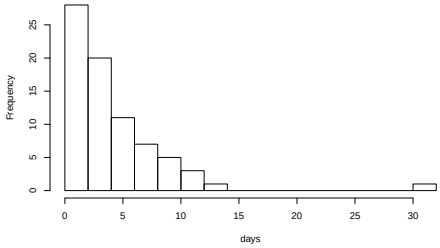
Before ER



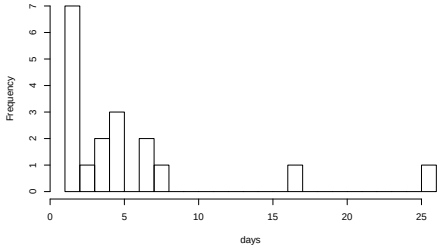
Before Hospitalized



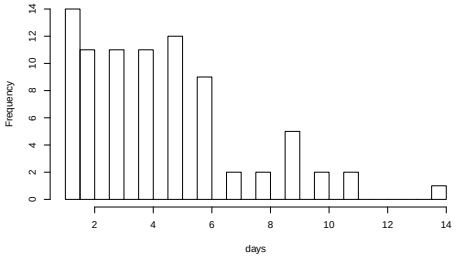
Before ICU



Before Ventilation



Before Intubation



Before Deceased

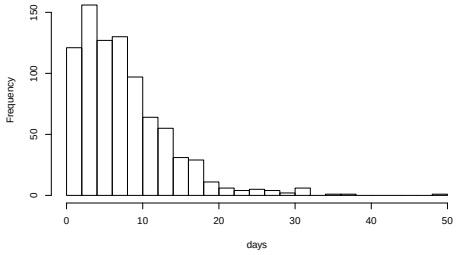


Figure 13: Durations of State D Prior to Transition

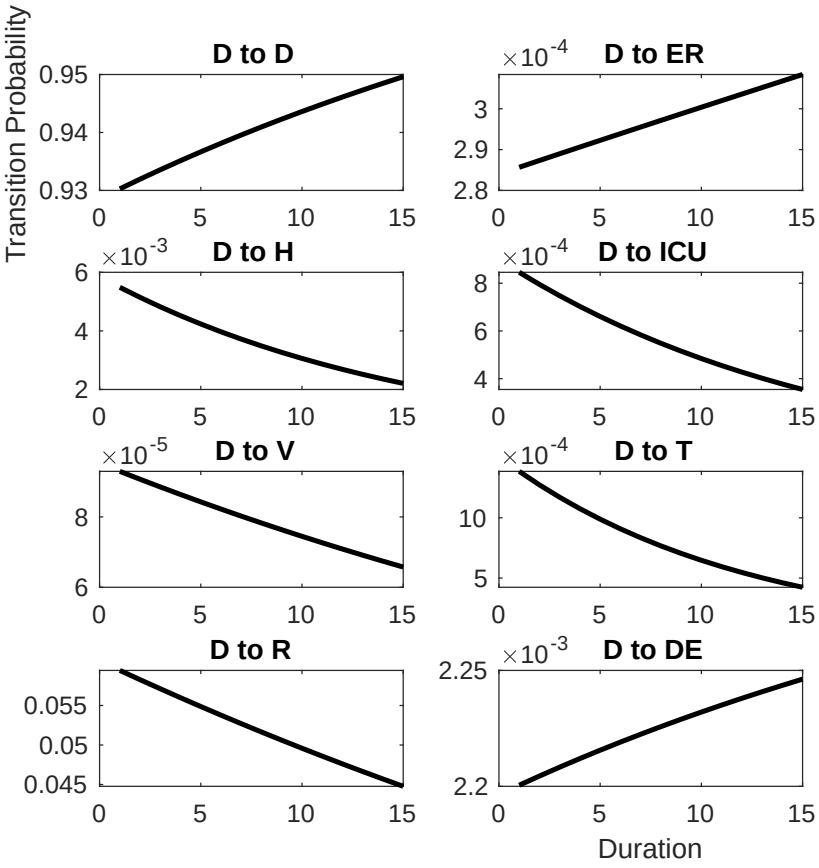


Figure 15: Transition Probabilities from "D" as Functions of Duration

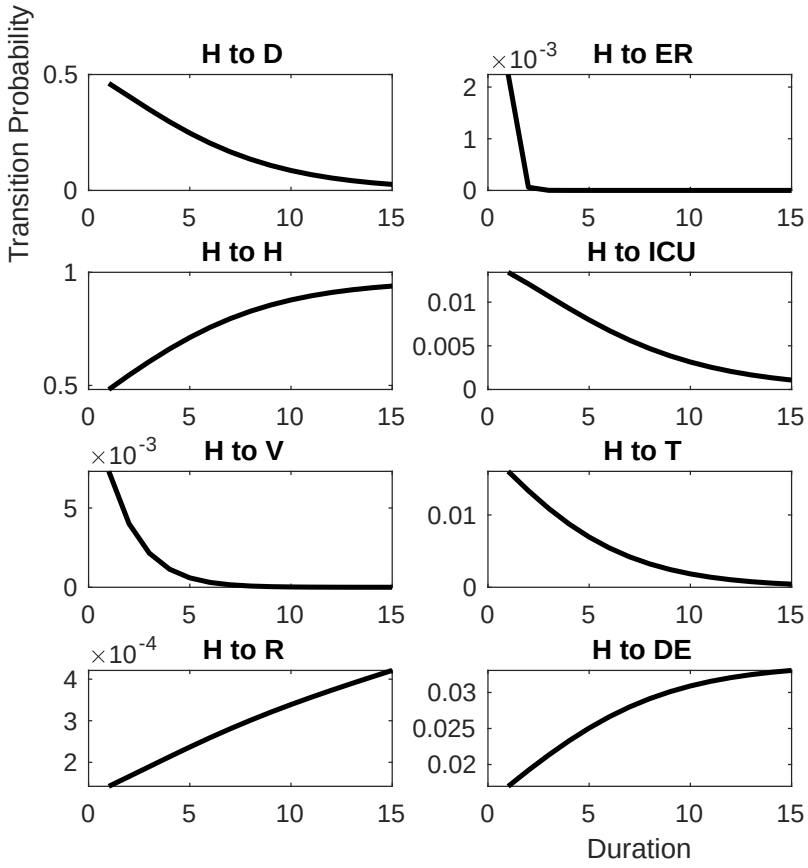


Figure 16: Transition Probabilities from "Hospitalized" as Functions of Duration

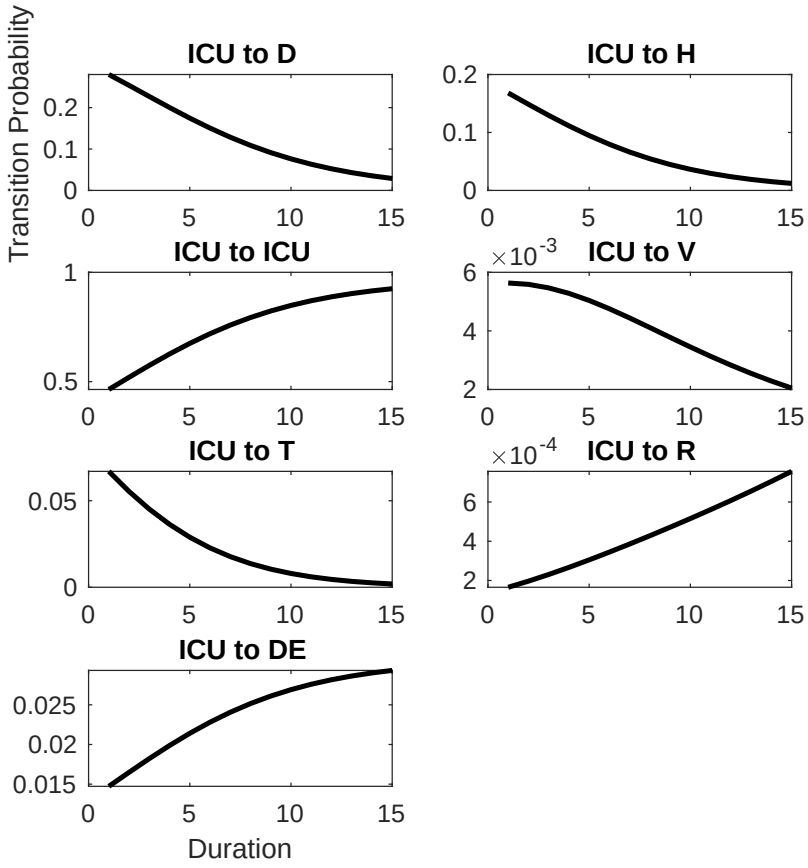


Figure 17: Transition Probabilities from "Intensive Care Unit" as Functions of Duration

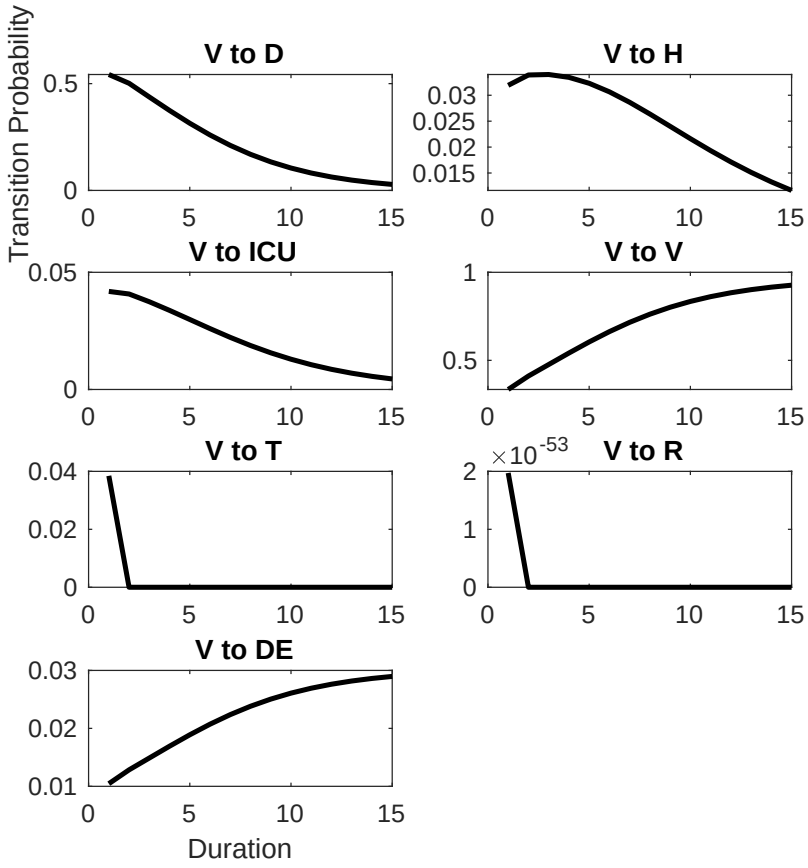


Figure 18: Transition Probabilities from "Ventilation" as Functions of Duration

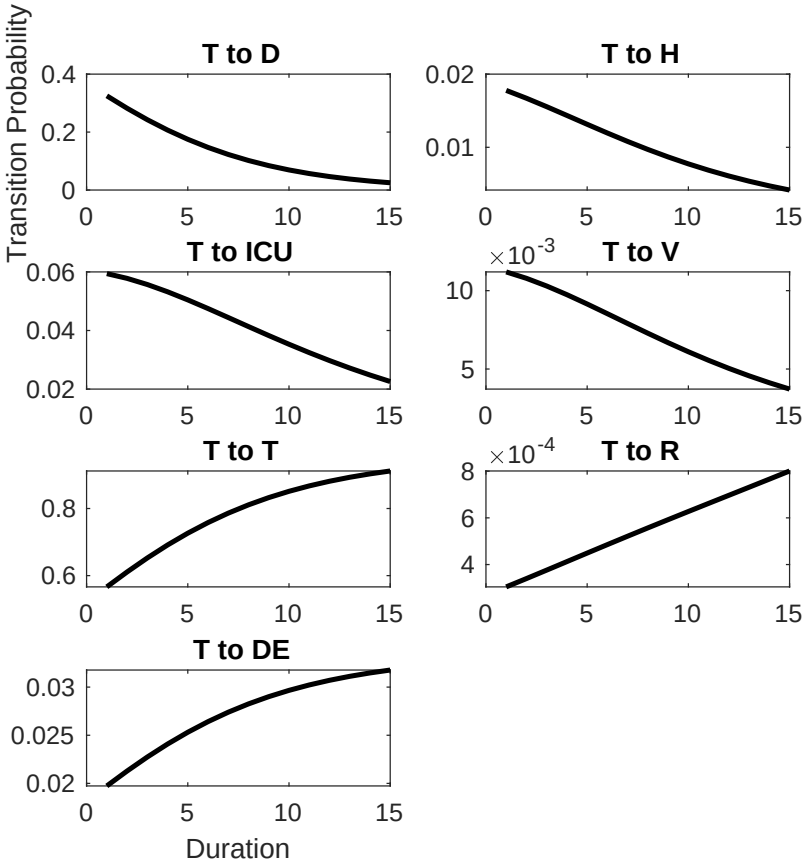


Figure 19: Transition Probabilities from "Intubation" as Functions of Duration